Improvement of a P-wave detector by a bivariate classification stage.

A.I. Hernández[1,2], G. Carrault[1], F. Mora[2]

[1]Laboratoire Traitement du Signal et de l'Image, INSERM,
Campus de Beaulieu, Université de Rennes I, 35042 Rennes Cedex France

[2] Grupo de Bioingeniería y Biofísica Aplicada,
Universidad Simón Bolívar, Apartado 89000. Caracas, Venezuela

Abstract.

For more than three decades, beat-to-beat detection of electrical atrial activity has been subject of interest in the biomedical field. Several detection schemes and algorithms have been proposed, however, this problem has not been satisfactory solved and remains as the main error source for accurate automatic detection and classification of supraventricular arrhythmias, mainly in ambulatory applications and in coronary care units. Some detection methods have been proposed, but none of them have been quantitatively tested or compared, due in part to the lack of a database with annotated *P*-waves. In this work, a classification stage is proposed as a new decision strategy to improve a *P*-wave detector previously developed in our laboratory. A quantitative evaluation of the original detector, the proposed classification stage and two other classical *P*-wave detectors is performed by means of the calculation of their receiver operating characteristics (ROC) curves. Results show the increased performance provided by the new classification stage with respect to the originally developed algorithm and two classical *P*-wave detection methods.

Keywords: *P*-wave detection, bivariate bayesian classification, ECG.

Introduction

Although the detection of the ventricular activity, reflected in the electrocardiogram as the *QRS* complex, has been widely refined during the last decades, the development of a robust method for beat-to-beat atrial activity detection is still one of the most challenging subjects in biomedical engineering. Several detection schemes and algorithms have been proposed, however, this problem has not been satisfactory solved and remains as the main error source for accurate automatic detection and classification of cardiac arrhythmias [1]. The problem becomes more complex when a low signal-to-noise ratio (SNR) is measured, which often occurs in situations where accurate detection is needed, such as in care units, during ambulatory recordings or implantable devices. Some of the most important complications involved in atrial activity detection are [1,2]:

- Low amplitude of the *P*-wave, which can also present important morphological variations.
- Lack of exclusive characteristics of the atrial electrical activity representation in the time or frequency domains. The bandwidth of the *P*-wave overlaps with the bandwidths of the *QRS* and *T*-wave, and has less energy.
- Abnormal atrial activations can occur simultaneously with the *QRS* or the *T*-wave, so a direct cancellation of ventricular complexes, attempting to improve the SNR, can cause the cancellation of atrial activity.

Due to the lack of an annotated *P*-wave database, previously proposed methods have been poorly evaluated, only by qualitative means, without providing any performance comparison with other methods.

In this paper, a general review of current approaches for *P*-wave detection is presented, including a detection scheme previously developed in our laboratory [3]. The presentation and evaluation of an enhancement for this detection scheme, based on an appended bivariate classification stage, is the main

purpose of this work. Basic theory of multivariate classification is also presented, leading to the proposed additional classification stage for *P*-wave detection.

Previous Atrial activity detection methods

Classical methods for atrial activity detection can be grouped as follows: *i)* Localized search, *ii)* Ventricular activity cancellation, *iii)* Direct detection methods, and *iv*) EECG-based methods.

Localized search methods

These are simple algorithms based on a localized search. In general, atrial activity detection is performed by identifying its associated *QRS* complex and searching backwards into a previously defined search window, normally taken from a fixed amount of samples before the detected *QRS*. Different combinations of algorithms have been presented in each step. Fixed-width [3,4] or variable-width [5] search windows, linear transformations for *P*-wave detection [3], time-frequency applications [5,6] or feature extraction and evaluation [4] have been proposed. Localized search methods for *P*-wave detection present acceptable performances under low-noise conditions and during sinus rhythm. The main problem of this type of algorithms is their dependency on the relative location of *P*-waves with respect to the *QRS*, making them unable to detect atrial activity in case of atrioventricular (AV) dissociation, or *QRS* miss-detections. Also these methods often generate false alarms when an artifact is detected as a *QRS* complex.

Ventricular activity cancellation

Ventricular activity cancellation methods have been widely used for *P*-wave detection from ECG signals. They involve a first stage of *QRS* cancellation, used to increase the SNR for a second stage of final event detection. Their usefulness has been centered on their ability to perform atrial activity detection in presence of AV dissociation. In general, these methods require the detection and cancellation of the *QRS* complex and *T*-wave, the application of some nonlinear transformation to the resultant signal in order to enhance atrial activity over the residual noise, and thresholding. Cancellation of the ventricular activity can be done directly or adaptively.

A classic algorithm for direct ventricular activity cancellation, based on the length transformation (LT), was introduced by Gritzali et al [7]. Atrial activity detection based in direct *QRS* cancellation methods has, as major drawback, its inability to detect overlapped *P*-waves on *QRS* complexes or *T*-waves.

An adaptive cancellation method was introduced by Thakor and Yi-Sheng [2]. Although its limitations in presence of sudden *QRS* morphology variations and miss detections of *QRS* complexes, this method provides cancellation of ventricular components without significantly distorting possible overlapped *P*-waves. Another adaptive method for ventricular activity cancellation has been recently proposed by Sornmo [4]. It is based in the substraction of an adaptive *QRS-T* template from the observed ECG signals after *QRS* detection and the application of a spatio-temporal alignment technique, which allows to reduce the amount of residual error due to the detection jitter and small morphological changements in the *QRS-T,* like those produced by a change in the electrical axis of the heart. However, this method remains very sensitive to the presence of pathological beats.

EECG-based detectors

One way to increase the SNR for atrial activity detection is to adapt the sensor to measure an appropriate representation of the activity of interest. The esophageal electrocardiogram (EECG) was introduced in this sense. This signal is acquired by means of a semi-invasive technique in which a small bipolar "pill-electrode", located in the esophagus, is used to obtain an amplified representation of the electrical atrial activity (*A*-wave) with respect to a reduced ventricular activity (*V*-wave) [8]. Detection of the *A*-wave is based on similar techniques as the *QRS* detection due to its morphological similarities. However, the morphology of EECG waves is very dependent on an appropriate location of the sensor, which can not be easily fixed in the appropriate place throughout the signal acquisition process.

Initial proposed detection scheme

A general structure for atrial activity detection was developed in our laboratory to improve the performance of current detection schemes. In this structure, multiple preprocessors having different sensors as input, feed a global fusion decision level for final event detection. An extended description of this general data fusion structure can be seen elsewhere [9,10].

Each preprocessor is in charge of all signal processing stages that can produce a valid statistic for atrial activity detection, from a given set of sensors. A preprocessor is composed of three levels: The first having a ventricular activity detection and cancellation based on an adaptive filter, the second uses another adaptive filter for *QRS* residual suppression, and the third produces a fiduciary mark of *P*-wave detection. Figure 1 shows a block diagram of the preprocessor.

Figure 1.

The length transformation, with parameters set to *QRS* detection, is used as the ventricular enhancement algorithm (D_1) applied to $S_{V,i}$. A threshold-based event detector (T_{si1}) is applied to the resulting transformation (S_1) to derive *QRS* detection impulses (i_1). The ventricular cancellation algorithm (AF_1) is an implementation of the *QRS-T* cancellation adaptive filter proposed by Thakor and Yi-Sheng, which uses i_1 as a reference signal to filter $S_{A+V,i}$, deriving an error signal (err_1) containing *P*-waves and cancellation residual noise.

A low pass FIR filter (F_1) was designed to enhance the *P*-wave and partially attenuate the high-frequency residue of *QRS* and *T* cancellation. The filtered signal is then squared. The resulting signal (S_2) is merged by means of another adaptive filter in which an index of the undesired ventricular activity (S_1) is used as reference. This level is conceived to eliminate remaining ventricular activity in S_2, taking into account the correlated information between S_1 and S_2. The resulting error signal ($T_{2,i}$) is processed by a band-pass filter (F_2). Threshold-based detection is then applied to obtain a local detection (u_i), used as input to the final data fusion stage.

Even if results reported in [9,10] were better than all the other algorithms tested, the number of false alarms, mainly on ECG channels with very low SNR and pathological beats, was still high due to the *QRS* residual and artifact energy. In order to reduce this false alarm rate, the threshold-based detection

module implemented originally for level 3 of each preprocessor was changed for a classification stage. The new implementation of level 3 (Figure 2) is based on a bivariate classification scheme, which consists to construct a description vector of dimension 2, each time a *P*-wave is suspected, and to classify it as a real *P*-wave, *QRS* residual, or noise. The method is detailed below.

Figure 2.

The multivariate classification stage.

At the output of the level 2 of each preprocessor, the final *P*-wave detection can be viewed as the classification of *P*-wave candidates into one of the following hypothesis: we observe only background noise (H_0), we observe background noise and high energy *QRS* residuals (H_1), or we observe a *P*-wave embedded in possible *QRS* residuals and background noise (H_2). We can state this classification problem as follows:

$$\begin{aligned} H_0 &: \; y_i(t) = n_i(t) \\ H_1 &: \; y_i(t) = A_i(t) + n_i(t) \\ H_2 &: \; y_i(t) = P_i(t - \tau_{i,j}) + A_i(t) + n_i(t) \end{aligned}$$

where $y_i(t)$ represents the transformation $T_{2,i}$ of the i^{th} observed sensor, $n_i(t)$ is a stationary gaussian noise, $A_i(t)$ is an impulsive noise, and $P_i(t-\tau_{i,j})$ is the j^{th} *P*-wave manifested on time $\tau_{i,j}$. Each element w from the set of all detected events Ω, is associated with one of the $l = 0...k$ hypothesis, by means of a decision ∂_l which is based on an observation vector of p different descriptive variables $X_w = (x_q, q = 1,...,p)$. Classification for a given X_w can be performed, using the Bayes approach, by selecting the decision ∂^* that minimizes the risk of bad classification:

$$R\partial^*(l) = \left\{ X_w \in \Re^p \,/\, \sum_{\substack{l'=0 \\ l' \neq l}}^{k} C_{l/l'} P(H_{l'}) f_{l'}(X_w) < \sum_{\substack{l'=0 \\ l' \neq m}}^{k} C_{m/l'} P(H_{l'}) f_{l'}(X_w), \forall m = 0,..,k, m \neq l \right\} \quad (1)$$

where $C_{l/l'}$ is the cost of choosing hypothesis H_l when $H_{l'}$ is true, $P(H_{l'})$ is the *a priori* probability of hypothesis *l'* and f_l is the theoretical probability distribution function of variable X_w. To apply the presented multivariate classification stage at the output of each preprocessor, an estimation of the costs and *a priori* probabilities of each hypothesis has to be made, appropriate descriptive variables have to be found, and the gathered data should fit some known probability distribution function.

In the case of three hypothesis ($k=2$) and equal costs of error $C_{l,l'} = 1 \quad \forall l \neq l'$, equation 1 becomes:

$$R\partial^*(l) = \left\{ X_w \in \Re^p \,/\, P(H_l) f_l(X_w) > P(H_m) f_m(X_w), \;\; \forall m = 0,k; \; m \neq l \right\} \quad (2).$$

Detecting a *P*-wave is equivalent to choose hypothesis H_2 in the classification stage. A proper classification rule for *P*-wave detection can be directly derived from equation 2. Detection of a *P*-wave will only occur when both of the following inequalities are satisfied:

$$\frac{f_2(X_w)}{f_0(X_w)} > \frac{P(H_0)}{P(H_2)} = s_1 \quad \text{and} \quad (3a)$$

$$\frac{f_2(X_w)}{f_1(X_w)} > \frac{P(H_1)}{P(H_2)} = s_2 \quad (3b)$$

The left-hand side of each inequality is the likelihood ratio of the corresponding hypothesis. Instead of trying to estimate the *a-priori* probabilities of each hypothesis, the right side of equations 3a and 3b can be considered as detection thresholds. Thus, the implementation of the classification rules depends on two thresholds, requiring a troublesome work for a proper estimation of its values. In order to avoid these problems, a sub-optimal classification rule has been constructed, depending only on one threshold.

Being both sides of the inequalities 3a and 3b always positive, each time they are satisfied, the following inequality will also be satisfied (although this is not reciprocal):

$$\frac{f_2(X_w)^2}{f_0(X_w)f_1(X_w)} > \frac{P(H_0)P(H_1)}{P(H_2)^2} \quad (4)$$

Even though equation 4 is a sub-optimal simplification of the detection rule presented in equations 3a and 3b that can produce some miss classifications (or false alarms), this new indicator has as advantage that it only depends on one comparison that can be implemented as a single threshold-based detection, leading to an easier detector evaluation and implementation.

Appropriate descriptive variables

A study of the properties of the transformed signal of each pre-processor ($T_{2,i}$) lead us to the following remarks:

- Mean peak amplitudes of $T_{2,i}$ in presence of H_0, H_1 and H_2 is different, being higher near H_2.
- Mean curvature of peaks associated with H_2 is lower than in other cases.

Polynomial fitting, centered at each transformation's peak, or *P*-wave candidate, could derive representative features of these quantities from the transformed signal, under each hypothesis. Let

$$P_w(k) = a_0 + a_1 k + a_2 k^2, \quad k = \tau_w - 10, \ldots, \tau_w - 10 \quad (5)$$

be this polynomial where τ_w is the time instant of the detected peak w (Figure 2). The bivariate descriptive variable $X_w = (x_1, x_2)^T$ is derived from the polynomial where x_1 is an estimation of the peak amplitude $P_w(\tau_w) = a_0$, and x_2 is an indication of the curvature of peak w *(i.e. a_2)*. It is important to emphasize that the derivation of these parameters does not depend on any threshold definition, as they are taken directly from each peak of the studied transformation, where the presence of a *P*-wave is suspected.

Selection of the probability distribution function

The 2-D distribution function of $X_w = (x_1, x_2)^T$, was determined for each hypothesis H_0, H_1 and H_2, applying the two-dimensional Kolmogorov-Smirnof (KS) test on some well known joint probability density functions. The KS tests were performed on two records *(100 and 108)* from the MIT-BIH database, in which *P*-waves were previously annotated. Testing sets were constructed with maximum peak values and coefficient a_2 for 300 randomly selected events in each evaluated record. The normality assumption could not be rejected in all the tests performed, having probabilities above 0.5 for

H_2 and above 0.3 for H_0 and H_1. Values of X_w used for the KS test were also used to estimate the two first statistical moments for each hypothesis, needed for the detector implementation.
Figure 3 shows the scatter plot of record 108, showing the clusters associated with each hypothesis.

Figure 3.

Assuming that $X_w=(a_{Ow}, a_{2,w})$ follows a Normal joint probability density function :

$$f_l(X) = \frac{1}{2\pi^{(N+M)/2}|C_l|^{1/2}} e^{\left(-\frac{1}{2}(X-m_l)^T C_l^{-1}(X-m_l)\right)} \tag{6}$$

where N=M=1, m_l is the mean vector and C_l is the covariance matrix estimated for hypothesis l, the logarithm of the detection rule presented in equation 4 becomes:

$$\log\left(\frac{|C_0||C_1|}{|C_2|^2}\right) + 2(X-m_2)^T C_2^{-1}(X-m_2) - \sum_{l=0}^{1}(X-m_l)^T C_l^{-1}(X-m_l) > 2\log\left(\frac{P(H_0)\,P(H_1)}{P(H_2)^2}\right) \tag{7}$$

Detection of a *P*-wave is acknowledged if the following condition is satisfied:

$$\log\left(\frac{|C_0||C_1|}{|C_2|^2}\right) + 2(X-m_2)^T C_2^{-1}(X-m_2) - \sum_{l=0}^{1}(X-m_l)^T C_l^{-1}(X-m_l) > s \tag{8}$$

where s is the threshold used to accept or reject the hypothesis H_2.
The presented detection rule is applied in level 3 of each preprocessor. A review of the whole *P*-wave detection procedure is presented in table 1.

Results.

Performances were evaluated by means of traditional Receiver Operational Characteristics curves (ROC), by plotting the number of false alarms (Nfa) versus the probability of detection (Pd). The evaluation of the detector using the original classification rule, presented in equations 3a and 3b, was also performed by estimating the surfaces for the probability of detection and number of false alarms generated by all combinations of the two thresholds within a predefined range. From these surfaces, a third surface was generated, to obtain an index of the probability of detection error. An optimal trace, consisting on the list of threshold pairs (s_1, s_2) that minimizes the probability of error, was found by applying a steepest descent algorithm. Corresponding values of probability of detection and number of false alarms, for each threshold pair in the list, were used to produce a ROC curve that is a function of the two thresholds. This curve, which will be called "optimal" for simplicity, is compared with the one produced by the simplified decision rule in order to measure its differences.
Figures 4a and 4b show *P*-wave detection results for records 100 and108 of the MIT-BIH database. They present ROC curves for the original detector, for the simplified classification rule, for the optimal classification rule, and for the algorithms proposed by Thakor [2] and Gritzali [7].

The algorithm proposed by Gritzali always provides the lowest performance. The curve presented by the algorithm proposed by Thakor is always close to the curve of the original detector, which can provide a reduced Nfa in signals presenting noise and *QRS* morphological variations (like in record 108) at the price of a reduced Pd for very high Nfas.

Minimal differences can be observed between the optimal and simplified version of the classification stage. Both classification rules provide the best performance for Nfas < 10^3. The difference in performance gain, provided by the classification stage, between both evaluated records may be associated to the reduced *P*-wave energy present in record 100 and the stability of the *QRS* morphology throughout this record, which permits a relatively good cancellation of the ventricular activity. These characteristics of record 100 also explain the reduced performance gain between the original detector and the detector of Thakor (see figure 4), whereas in record 108, a higher performance gain can be measured between these detectors.

Two points are marked over the ROC curves of the optimal and simplified detection rules. The pentagon represents the (*Pd, Nfa*) couple corresponding to the theoretical threshold, calculated by resolving the right-hand side of equation 7, and thus providing the minimal probability of error. The estimation of the probabilities of each hypothesis, to calculate this theoretical threshold, was made using available *P*-wave and *QRS* annotations. The diamond represents the (*Pd, Nfa*) couple which provides the minimal probability of error, obtained experimentally. It is important to notice the closeness of both marks, indicating that an appropriate threshold definition for the application of the proposed detector can be obtained if the values of the probabilities of each hypothesis can be estimated.

Figure 4.

Discussion

In this work, a bivariate classification stage was designed to replace the threshold-based decision rule used in our initial *P*-wave detection algorithm. This classification stage leads to a reduced number of false alarms and improved robustness, when compared with our original detector and two other detectors presented in the literature. Two different implementations of this classification scheme, based on a Bayes approach, were presented and quantitatively evaluated: an "optimal" implementation, derived directly from the theory, and a simplified version, which is easier to implement. Although some miss classifications can be obtained from the application of the simplified decision rule, the results present no significant differences between both rules.

The application of the proposed classification scheme requires the previous estimation of the two first statistical moments associated with each hypothesis. An optimal value for *s*, which minimizes the probability of error, can be also estimated from the *a priori* probabilities of occurrence of events in each hypothesis $P(H_0)$, $P(H_1)$ and $P(H_2)$. Although this kind of estimation can be difficult to implement in real applications, the ability to offer a possible way to estimate an appropriate operational threshold is one of the important advantages of the proposed detector over most detection schemes, in which threshold estimation can become a complex problem.

The classification stage proposed in this paper, both in its "optimal" or simplified versions, leads to an improved detection performance, by reducing the probability of false alarm, up to an order of 10 (see figures 4a and 4b) with respect to the original algorithm and the one proposed by Thakor. However, the development of a robust algorithm for atrial activity detection, to be used in clinical applications, is still a major need, and in the authors' opinion, could not be based only on ECG signals. Results of this work make us think that a possible way to overcome the problems associated with atrial activity detection should be based on data fusion techniques, which can properly integrate information from different sources and sensors, as reported in [10].

References

[1] J. Jenkins, "Automated Electrocardiography and Arrhythmia Monitoring," *Prog. Card. Dis.* vol. 25. no. 5, pp. 367-408, 1983.

[2] N. Thakor, Z. Yi-Sheng. "Applications of Adaptive Filtering to ECG Analysis: Noise Cancellation and Arrhythmia Detection," *IEEE Trans. Biom. Eng*, vol. 38, no. 8, pp. 785-794, 1991.

[3] Freeman K, Singh A. P Wave detection of ambulatory ECG. In: Proc. Ann. Int. Conf. IEEE/EMBS. vol. 13. 1991, pp. 647-648.

[4] Reddy S, Elko P, Chistenson D, Rowlandson GI. Detection of P waves in resting ECG: A Preliminary Study. In: Computers in Cardiology 1992. Los Alamitos: IEEE Computer Society Press, Sept 1992, pp. 87-90.

[5] Fokapu O, Girard JP. A New Approach for P Wave Detection Using Analytic Signal. In: Proc. Ann. Int. Conf. IEEE/EMBS. 1993, pp. 400-401.

[6] Li C, Zheng C, Tai C. Detection of ECG Characteristic Points Using Wavelet Transform. *IEEE Trans. Biom. Eng*. vol 42, no 1, pp. 21-28, 1995.

[7] F. Gritzali, G. Frangakis, G. Papakonstantinou. "Detection of the P and T waves in an ECG". *Computers and Biomedical Research* ; vol 22, pp. 83-9, 1989.

[8] R. Arzbaecher, "A pill electrode for the study of cardiac arrhythmia," *Med Instrum*, vol. 12, pp. 277-81, 1978.

[9] A. Hernández, G. Carrault, F. Mora, G. Passariello, M.I. Hernández, J. Schleich. "Quantitative Comparison of Atrial Activity Detection Approaches". In: Computers in Cardiology 1997. Los Alamitos: IEEE Computer Society Press, 1997, pp. 481-484.s

[10] A. Hernández, G. Carrault, F. Mora, G. Passariello, J. Schleich. "Multisensor Fusion for Atrial and Ventricular Activity Detection in Coronary Care Monitoring". Paper accepted *IEEE Trans. Biom. Eng*. Special Issue on Data Fusion.

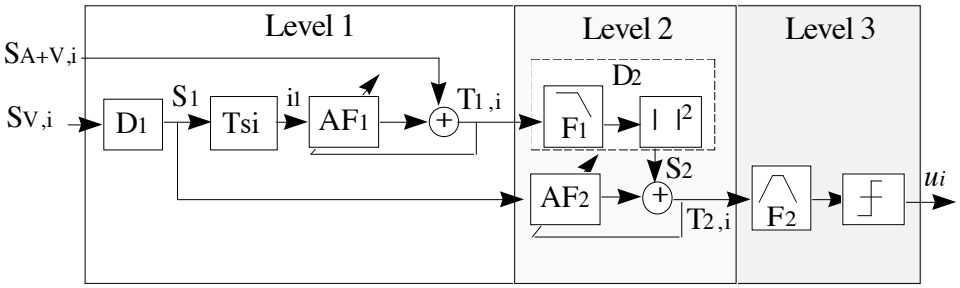

Figure 1. Preprocessor implementation.

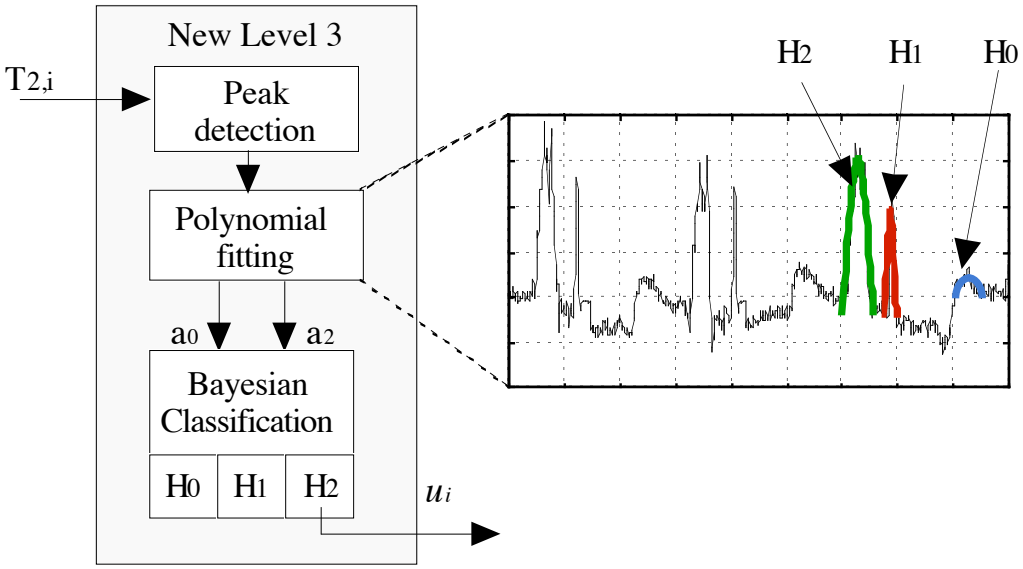

Figure 2. Multivariate classification stage for the new level 3. The gray line marked with H_2 is an example of polynomial fitting on a candidate *P*-wave (see text for details).

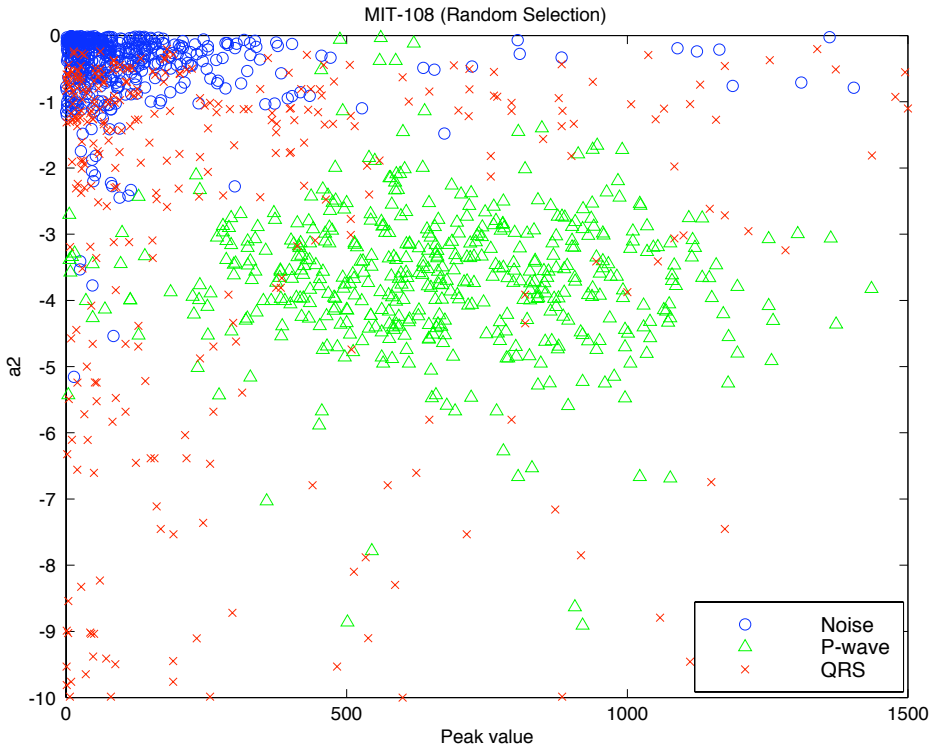

Figure 3. Scatter plot of peak value (a_0) and coefficient a_2 for 300 *P*-wave candidates taken randomly from register 108 of the MIT-BIH DB, falling into the three defined hypothesis (o- H_0, x- H_1, Δ- H_2).

Table 1. Detector implementation.

1. Application of preprocessor i, $i=1\ldots N$, $N>0$ as described in section 3.
2. For each candidate P-wave detected at the output of each preprocessor ($T_{2,i}$):
   - A quadratic polynomial is fitted in the LMS sense.
   - A vector $X=(a_0,a_2)^T$ of descriptive variables is constructed from polynomial coefficients.
   - The following detection rule is applied:

$$\frac{f_2(X_w)^2}{f_0(X_w)f_1(X_w)} > \frac{P(H_0)P(H_1)}{P(H_2)^2}$$

   using previously estimated values for the two first statistical moments of each hypothesis, to generate the local P-wave detection for preprocessor i (u_i).
3. If more than one preprocessor has being used ($N>1$), all local detections u_i are used by a global data fusion module, in order to produce the final P-wave detection.

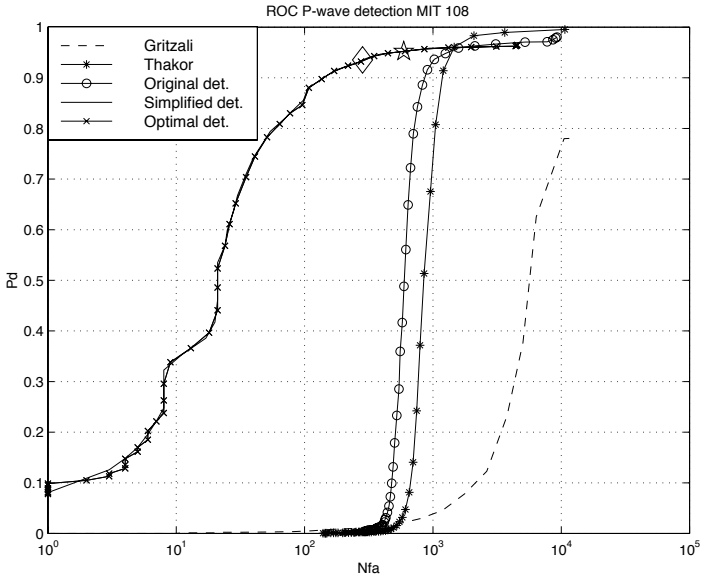

Figure 4a. ROC curve for record 108: The diamond represents the experimental optimal threshold and the star represents the theoretical threshold.

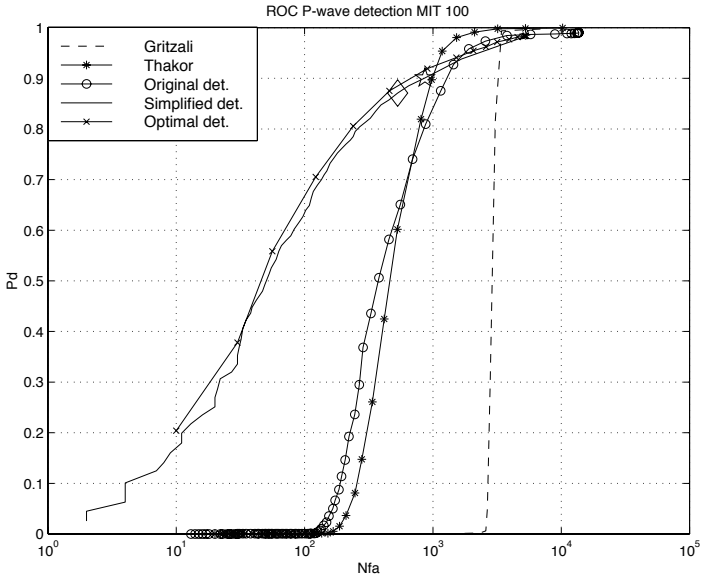

Figure 4b. ROC curve for record 100: The diamond represents the experimental optimal threshold and the star represents the theoretical threshold.